Consensus in the Wasserstein Metric Space of Probability Measures[*]

Adrian N. Bishop [†]   Arnaud Doucet [‡]

Abstract

Distributed consensus in the Wasserstein metric space of probability measures is introduced in this work. Convergence of each agent's measure to a common measure value is proven under a weak network connectivity condition. The common measure reached at each agent is one minimizing a weighted sum of its Wasserstein distance to all initial agent measures. This measure is known as the Wasserstein barycentre. Special cases involving Gaussian measures, empirical measures, and time-invariant network topologies are considered, where convergence rates and average-consensus results are given. This algorithm has potential applicability in computer vision, machine learning and distributed estimation, etc.

1 Introduction

The problem of distributed consensus concerns a group of agents that seek to reach agreement upon certain state variables of interest by exchanging information across a network. Typically the agents are connected via a network that changes over time due to link failures, node failures, packet drops etc. For example, in distributed sensor networks the interaction topology may change over time as individual nodes (or some subset of such) may be mobile or unreliable or communication constraints are also present. All such variations in topology can happen randomly and often the network is disconnected for some time. Studies on the convergence of consensus algorithms (to a common agreed 'value' at each agent) are often motivated by such complex time-varying networks.

1.1 Background

The consensus problem has a long history, e.g [1], which is too broad to cover here. We highlight [2–7] for further history and background.

Many consensus algorithms have been proposed in the literature. References [2,4,6,8] focus on linear update rules (at each agent) and typically concern average-consensus or consensus about some linear function of all initial agent states in Euclidean space. The average-consensus problem has a natural relationship with distributed linear least squares or distributed (linear) maximum likelihood estimation [9] and distributed Kalman filtering [10–13]. Alternative consensus algorithms using nonlinear update rules have been proposed and studied in [5,14–16]. Here consensus to general functions (e.g. the maximum or minimum etc.) of all initial agent states may be sought as in [17,18] and even finite-time convergence may be achievable [18,19]. One may also want to achieve consensus to some time-varying reference signal as in [20,21].

We note here that the majority of the literature on consensus concerns agreement in Euclidean space as exemplified by the seminal papers of [2–7]. However, there are exceptions. The problem of synchronisation is closely related to consensus but typically deals with the problem of driving a network of oscillators to a common frequency/phase. This work typically concerns nonlinear manifolds such as the circle. A survey on synchronisation is given in [22] while consensus and synchronisation are related in [23]. Some other notable exceptions of consensus in non-Euclidean spaces are [24–29]. In particular, [25, 27] consider general nonlinear consensus on manifolds by embedding such manifolds in a suitable high-dimensional Euclidean space. In particular, this embedding approach is used to perform consensus on the special orthogonal group and on Grassmann manifolds. The authors in [24, 26, 28] study consensus in different metric spaces which is more closely related to the present work. For example, the author of [28] develops an analogue of Wolfowitz's theorem [3] for a class of metric spaces with non-positive curvature which leads to a notion of consensus in such spaces.

1.2   Contributions

The main contributions of this paper are a novel algorithm and convergence results for distributed consensus in the space of probability measures with time-varying interaction networks. We introduce a well-studied metric known as the Wasserstein distance which allows us to consider an important set of probability measures as a metric space [30]. The proposed consensus algorithm is based on iteratively updating each agent's probability measure by finding a measure that minimizes the weighted sum of its Wasserstein distances to the agent's own previous measure plus all neighbour agents' measures. We show that convergence of the individual agents measures to a common probability measure is guaranteed under a weak network connectivity condition. The common measure that is achieved asymptotically at each agent is the one that is closest simultaneously to all initial agent measures in the sense of the Wasserstein distance.

This work has potential applicability in the field of distributed estimation, distributed information fusion and machine learning among other fields discussed later. For example, suppose each agent starts with a (posterior) probability measure associated to some common underlying event of interest. Then one may like to combine all these measures (which amount to each agents estimate and/or belief of the underlying event) into a common probability measure that captures all the agents beliefs. The proposed consensus algorithm can achieve this in a very general distributed setting. Related work in [31] considers the application of consensus to the problem of distributed Bayesian computations. In [32] the consensus algorithm from [31] is further studied and applied in distributed estimation. Unlike [31], the proposed method here can deal with singular measures, may be robust to unknown correlations, and the required connectivity condition is weaker [33, 34]. We note that a Monte Carlo approximation of the consensus algorithm from [31] was also studied in [35]. In [29] a consensus algorithm on the Riemannian manifold of (Gaussian) covariance matrices is introduced under the Fisher metric (related to to the Kullback-Leibler distance).

Further discussion on applications and related topics, particularly relevant in the Wasserstein domain, is given later.

This paper extends [36] with the addition of results detailing convergence rates and network properties in which particular consensus values may be achieved.

1.3   Paper Organization

The main contribution is given in Section 2 where a consensus algorithm in the space of probability measures is introduced, and its convergence is studied. In Section 3 we establish for specific scenarios - initial Gaussian measures; for empirical measures; and time-invariant networks - the exponential convergence of this consensus algorithm and various computational aspects. In Section 4 we discuss potential applications. Concluding remarks are given in Section 5.

1.4   Notation and Conventions

Consider a group of agents indexed in $\mathcal{V} = \{1, \ldots, n\}$ and a set of (possibly) time-varying undirected links $\mathcal{E}(t) \subset \mathcal{V} \times \mathcal{V}$ defining a network graph $\mathcal{G}(t)(\mathcal{V}, \mathcal{E}(t))$. The neighbour set at agent i is denoted by $\mathcal{N}_i(t) = \{j \in \mathcal{V} : (i, j) \in \mathcal{E}(t)\}$. Time is indexed using $\mathbb{N}$.

The graph adjacency matrix $\mathbf{A}(t) \in \mathbb{R}^{n \times n}$ obeys $\mathbf{A}(t) = \mathbf{A}(t)^\top = [a_{ij}(t)]$ where $a_{ij}(t) = 1 \Leftrightarrow (i, j) \in \mathcal{E}(t)$ and $a_{ij}(t) = 0$ otherwise. Implicit throughout is that $a_{ii}(t) = 1$ for all i and t and thus $i \in \mathcal{N}_i(t)$ for all t. A weighted adjacency matrix is denoted by $\mathbf{W}(t) = [w_{ij}(t)] \in \mathbb{R}^{n \times n}$ with $a_{ij}(t) = 1 \Leftrightarrow w_{ij}(t) > 0$ and $w_{ij} = 0$ otherwise. We require $\sum_{j \in \mathcal{N}_i(t)} w_{ij}(t) = 1$ and it holds that $w_{ii}(t) > 0$ for all i and t so that $w_{ij}(t) \in [0, 1)$ whenever $i \neq j$ and for all t.

The adjacency matrix $\mathbf{A}(t)$ defines $\mathcal{G}(t)(\mathcal{V}, \mathcal{E}(t))$ and vice versa because $a_{ij}(t) = 1 \Leftrightarrow (i, j) \in \mathcal{E}(t)$ and $a_{ij}(t) = 0 \Leftrightarrow (i, j) \notin \mathcal{E}(t)$. The weighted adjacency matrix defines $\mathcal{G}(t)(\mathcal{V}, \mathcal{E}(t))$ because $w_{ij}(t) > 0 \Leftrightarrow (i, j) \in \mathcal{E}(t)$ and $w_{ij}(t) = 0 \Leftrightarrow (i, j) \notin \mathcal{E}(t)$, but $\mathcal{G}(t)(\mathcal{V}, \mathcal{E}(t))$ alone defines only the sparsity pattern of $\mathbf{W}(t)$.

Consider the sequence of graphs $\mathcal{G}(t_k), \mathcal{G}(t_{k+1}), \ldots, \mathcal{G}(t_{k+T})$ on the same vertex set $\mathcal{V}$. The union of this sequence is denoted by $\mathfrak{G}(t_k, t_{k+1})(\mathcal{V}, \cup_{t \in [t_k, t_{k+T}]} \mathcal{E}(t))$, i.e. $\mathfrak{G}(t_k, t_{k+1})$ is just a graph on the vertex set $\mathcal{V}$ with edges $\cup_{t \in [t_k, t_{k+T}]} \mathcal{E}(t)$. The sequence is said to be jointly connected if $\mathfrak{G}$ is connected.

2   Consensus in the Wasserstein Space of Probability Measures

The main contribution of this work is given in this section where we introduce and establish the convergence of a consensus algorithm in the Wasserstein metric space of probability measures.

Suppose the state of agent i is given by a Radon probability measure μ_i defined on the Borel sets of $(\mathbb{R}, d)$ where in this section we restrict $d : \mathbb{R} \times \mathbb{R} \to [0, \infty)$ to be the usual Euclidean distance. Define the space of all such measures on $(\mathbb{R}, d)$ by $\mathfrak{U}(\mathbb{R})$ and the subset of all such measures with finite p^{th} moment by $\mathfrak{U}_p(\mathbb{R})$ where henceforth we assume that $2 \leq p < \infty$. That is, $\mathfrak{U}_p$ is the collection of probability measures such that $\int_{\mathbb{R}} d(x, x_0)^p \, d\mu_i(x) < \infty$ for a given, arbitrary, $x_0 \in \mathbb{R}$.

One can associate the Wasserstein metric $\ell_p : \mathfrak{U}_p(\mathbb{R}) \times \mathfrak{U}_p(\mathbb{R}) \to [0, \infty)$ with $\mathfrak{U}_p$ which is defined by

$$\ell_p(\mu_i, \mu_j) = \left(\inf_{\gamma \in \Gamma(\mu_i, \mu_j)} \int_{\mathbb{R} \times \mathbb{R}} d(x_i, x_j)^p \, d\gamma(x_i, x_j) \right)^{1/p}$$

where $\Gamma(\mu_i, \mu_j)$ denotes the collection of all probability measures on $\mathbb{R} \times \mathbb{R}$ with marginals μ_i and μ_j on the first and second factors; see [37, 38].

Let us recall some standard results about the Wasserstein metric space $(\mathfrak{U}_p(\mathbb{R}), \ell_p)$ when $p \geq 2$; see e.g. [37–40].

1. $(\mathfrak{U}_p(\mathbb{R}), \ell_p)$ is a complete and separable metric space.

2. $\lim_{k \to \infty} \ell_p(\mu_k, \mu) = 0$ is equivalent to weak convergence and convergence of the first p moments.

3. Given two measures $\mu_i, \mu_j \in \mathfrak{U}_p(\mathbb{R})$ then $\ell_p(\mu_i, \mu_j) = \ell_p(\mu_i, \mu) + \ell_p(\mu_j, \mu)$ for some $\mu \in \mathfrak{U}_p(\mathbb{R})$.

4. More generally, there exists a continuously parameterised constant speed path $\mu_s \in \mathfrak{U}_p(\mathbb{R})$, $s \in [0, 1]$ such that for $\mu_i, \mu_j \in \mathfrak{U}_p$ we have $\mu_{s=0} = \mu_i$ and $\mu_{s=1} = \mu_j$ and $\ell_p(\mu_i, \mu_j) = \ell_p(\mu_i, \mu_s) + \ell_p(\mu_j, \mu_s), \forall s \in [0, 1]$. The measure μ_s is known as the interpolant measure [41].

5. The interpolant measure defines a geodesic and consequently $(\mathfrak{U}_p, \ell_p)$ is geodesic.

6. $(\mathfrak{U}_p(\mathbb{R}), \ell_p)$ has vanishing curvature in the sense of Alexandrov (a subset of CAT(0)); see Proposition 4.1 in [39].

7. $(\mathfrak{U}_p(\mathbb{R}), \ell_p)$ is simply connected; see [39, 40].

All metrics are continuous and we recall that a constant speed geodesic in $(\mathfrak{U}_p(\mathbb{R}), \ell_p)$ is a curve $\mu_s : \mathbb{I} \to \mathfrak{U}_p$ parameterised on some interval $\mathbb{I} \subset \mathbb{R}$ that satisfies $\ell_p(\mu_{s_i}, \mu_{s_j}) = v|s_i - s_j|$ for some constant $v > 0$ and for all $s_i, s_j \in \mathbb{I}$.

Suppose the measure at agent i is updated by

$$\mu_i(t+1) = \operatorname*{argmin}_{\eta \in \mathfrak{U}_p(\mathbb{R})} \sum_{j \in \mathcal{N}_i(t)} w_{ij}(t) \ell_p(\eta, \mu_j(t))^p \tag{1}$$

for all $i \in \mathcal{V}$ where again we require $\sum_{j \in \mathcal{N}_i(t)} w_{ij}(t) = 1$ and $w_{ii}(t) > 0$ so that consequently $w_{ij}(t) \in [0, 1)$ whenever $i \neq j$. This operation is well-defined as discussed below.

Application of the update rule (1) to each agent $i \in \mathcal{V}$ corresponds to the proposed nonlinear (distributed) consensus algorithm.

2.1 Main Result

We state here our main result.

Theorem 1. *Consider a group of agents $\mathcal{V}$ and network $\mathcal{G}(t)(\mathcal{V}, \mathcal{E}(t))$ where each agent i has initial state $\mu_i(0) \in \mathfrak{U}_p(\mathbb{R})$ and updates its state $\mu_i(t) \in \mathfrak{U}_p(\mathbb{R})$ according to (1). If for all $t_0 \in \mathbb{N}$ the graph union $\mathfrak{G}(t_0, \infty)$ is connected then there exists $\mu^* \in \mathfrak{U}_p(\mathbb{R})$ such that*

$$\lim_{t \to \infty} \ell_p(\mu_i(t), \mu^*) = 0$$

for any $i \in \mathcal{V}$.

The proof of Theorem 1 is given in the next subsection following the provision of a number of supporting results.

2.2 Proof of the Main Result

Note that a subset $\mathfrak{X} \subset \mathfrak{U}_p(\mathbb{R})$ is convex if every geodesic segment whose endpoints are in $\mathfrak{X}$ lies entirely in $\mathfrak{X}$. The (closed) convex hull co($\mathfrak{Y}$) of a subset $\mathfrak{Y} \subset \mathfrak{U}_p$ is the intersection of all (closed) convex subsets of $\mathfrak{U}_p$ that contain $\mathfrak{Y}$.

Lemma 1. *If $\mu_i(t) \in \mathfrak{U}_p(\mathbb{R})$ then the operation (1) is well-defined in the sense that it admits a solution and this solution is unique whenever (at least) one $\mu_j(t) \in \mathfrak{U}_p$, $j \in \mathcal{N}_i(t)$ does not give support to small sets[1].*

Recall that $(\mathfrak{U}_p(\mathbb{R}), \ell_p)$ is CAT(0) (indeed it has the stronger property of vanishing curvature), in addition to being uniquely geodesic, complete and separable, i.e. $(\mathfrak{U}_p(\mathbb{R}), \ell_p)$ is a Hadamard space. This lemma then follows from the fact that $(\mathfrak{U}_p(\mathbb{R}), \ell_p)$ is Hadamard and Fréchet averages such as defined by operations of the form (1) are well defined in such spaces; see page 334 in [45]. The existence and uniqueness of solutions to (1) is also discussed in [42]

---

[1]A small set is defined [42] as a set of Hausdorff dimension 0. This condition plays a role only in uniqueness and it is generally unnecessary [42, 43]. However, this requirement does exclude empirical measures on $\mathbb{R}$ which arise in numerous applications relevant to this work (as discussed later). Luckily, it is generically true (i.e. excluding particular, non-generic, arrangements) that (1) has a unique solution even in such cases; see [43, 44]. Note if all inputs are discrete we allow for both common and uncommon supports. Going forward we will not repeatedly call on the need for (at least) one initial measure to exclude support on small sets and later results may be read as implicitly assuming uniqueness (or implicitly assuming exclusion of support on small sets).

more generally; see also [43, 44]. It is worth noting in passing the related work in [24, 28] which deals with similar consensus topics in CAT(0) spaces, and [26] which deals with consensus in a general class of convex metric spaces.

The convex hull of the set of measures $\{\mu_i\}$, $i \in \widetilde{\mathcal{V}} \subseteq \mathcal{V}$, is defined by

$$\mathrm{co}(\{\mu_i\}) = \{\operatorname*{argmin}_{\eta \in \mathfrak{U}_p(\mathbb{R})} \sum_{i \in \widetilde{\mathcal{V}}} w_i \ell_p(\eta, \mu_i)^p | w_i \geq 0, \sum_i w_i = 1\}.$$

Lemma 2. *Consider a collection $\{\mu_i\}$, $i \in \widetilde{\mathcal{V}} \subseteq \mathcal{V}$ of distinct measures in $(\mathfrak{U}_p(\mathbb{R}), \ell_p)$. The convex hull of $\{\mu_i\}$ is $\mathrm{co}(\{\mu_i\}) \subset \mathfrak{U}_p(\mathbb{R})$ and is isometric to a l-sided convex polygon in $\mathbb{R}^2$ with $2 \leq l \leq |\{\mu_i\}|$.*

Before proceeding with the proof we point to [46] for background on comparison triangles and Alexandrov curvature of metric spaces. We also note that in a general geodesic CAT(0) space, i.e. some arbitrary geodesic space with non-positive curvature, the preceding lemma is not true and the convex hull of a 'geodesic triangle' defined by three points in such spaces may be of dimension greater than two[2]; see Chapter II.2 in [46].

Proof. Lemma 2 is a simple consequence of the vanishing curvature property of $(\mathfrak{U}_p(\mathbb{R}), \ell_p)$. We elaborate for completeness. $(\mathfrak{U}_p(\mathbb{R}), \ell_p)$ has vanishing curvature in the sense of Alexandrov (see Proposition 4.1 in [39]) which formally means that for any triangle of points $\{\mu_i\}$, $i \in \{i_1, i_2, i_3\}$ and any point on the geodesic $\mu_s \in \mathfrak{U}_p(\mathbb{R})$, $s \in [0, 1]$ such that, for example, $\mu_{s=0} = \mu_{i_1}$ and $\mu_{s=1} = \mu_{i_2}$ then the ℓ_p distance between μ_{i_3} and μ_s, $s \in [0, 1]$ is the same as the corresponding Euclidean distance in a comparison triangle in $\mathbb{R}^2$. Consider also any pair of points μ_j and μ_k with μ_j on the geodesic connecting μ_{i_1} and μ_{i_2} and μ_k on the geodesic connecting μ_{i_1} and μ_{i_3} with $\{\mu_j, \mu_k\} \cap \{\mu_i\} = \emptyset$, $i \in \{i_1, i_2, i_3\}$. Then vanishing curvature also implies $\angle_{\mu_{i_1}}(\mu_j, \mu_k)$ is equal to the usual interior Euclidean angle at the corresponding vertex in the comparison triangle in $\mathbb{R}^2$. Here the angle $\angle_{\mu_{i_1}}(\mu_j, \mu_k)$ is the Alexandrov angle in arbitrary metric spaces; see Chapter II.1 in [46]. It now follows that the convex hull of any triangle of points $\{\mu_i\}$ in $(\mathfrak{U}_p(\mathbb{R}), \ell_p)$ is isometric to a triangle in $\mathbb{R}^2$; e.g. see Proposition 2.9 (Flat Triangle Lemma) in [46]. Now define $\mathfrak{C} = \{\Delta_j\}$ to be the collection of geodesic triangles in $(\mathfrak{U}_p(\mathbb{R}), \ell_p)$ defined by every combination of three points in $\{\mu_i\}$, $i \in \widetilde{\mathcal{V}} \subseteq \mathcal{V}$. Clearly $\mathrm{co}(\{\mu_i\}) = \cup_j \Delta_j$. Consider also the corresponding collection $\mathfrak{C}^+ = \{\Delta_j^+\}$ of comparison triangles in $\mathbb{R}^2$. The Flat Triangle Lemma implies that this collection can be arranged in $\mathbb{R}^2$ such that each angle $\angle_{\mu_i}(\mu_j, \mu_k)$ and each distance $\ell_p(\mu_i, \mu_j)$ for all $i, j, k \in \widetilde{\mathcal{V}}$ in $(\mathfrak{U}_p(\mathbb{R}), \ell_p)$ equals exactly the corresponding angle or distance in the comparison configuration of points in $\mathbb{R}^2$. Obviously, the convex hull of the comparison configuration is a l-sided convex polygon in $\mathbb{R}^2$ with $2 \leq l \leq |\{\mu_i\}|$ and equal to $\cup_j \Delta_j^+$. Define the following map

$$f_{\ell_p, d} : \mathrm{co}(\{\mu_i\}) \to \mathbb{R}^2, \quad i \in \widetilde{\mathcal{V}} \tag{2}$$

so the restriction

$$\begin{aligned} f_{\ell_p, d}(\Delta_j) &= f_{\ell_p, d}(\mathrm{co}(\{\mu_{j_1}, \mu_{j_2}, \mu_{j_3}\})) \\ &= \mathrm{co}(\{f_{\ell_p, d}(\mu_{j_1}), f_{\ell_p, d}(\mu_{j_2}), f_{\ell_p, d}(\mu_{j_3})\}) = \Delta_j^+ \end{aligned}$$

$\forall j \in \mathfrak{C} = \{\Delta_j\}$ is an isometry. Then

$$f_{\ell_p, d}(\mathrm{co}(\{\mu_i\})) = f_{\ell_p, d}(\cup_j \Delta_j) = \cup_j f_{\ell_p, d}(\Delta_j) = \cup_j \Delta_j^+$$

from the Flat Triangle Lemma and the property of vanishing curvature. For any two points in $\mathrm{co}(\{\mu_i\})$ there exists a $\Delta_j \in \mathfrak{C}$ that contains them and the restriction $f_{\ell_p, d}(\Delta_j)$ is an isometry to a convex subset of $\cup_j \Delta_j^+$. Thus, $f_{\ell_p, d}$ is an isometry and this completes the proof. $\square$

---

[2]Our Euclidean intuition is generally wrong when it suggests the existence of a two-dimensional convex hull for a triangle defined by three points and the geodesics connecting them (albeit this is hard to visualise of course).

Lemma 3. *Consider the convex hull* $\mathrm{co}(\{\mu_j(t)\})$*, with* $j \in \mathcal{N}_i(t)$ *at time* t*. If agent* i *applies* *(1) it follows that* $\mu_i(t+1)$ *is strictly within the convex hull* $\mathrm{co}(\{\mu_j(t)\})$ *whenever* $|\{\mu_j(t)\}| \geq 2$ *and two agent states are distinct and* $w_{ij}(t) \in (0,1)$*.*

Proof. It is enough to consider two agents [3] $i, j \in \mathcal{V}$ with (1) then given by

$$\mu_i(t+1) \quad = \quad \underset{\eta \in \mathfrak{U}_p(\mathbb{R})}{\mathrm{argmin}} \ w_{ii}(t) \left(\ell_p(\eta, \mu_i(t))^p - \ell_p(\eta, \mu_j(t))^p\right) + \ \ell_p(\eta, \mu_j(t))^p$$

and to note that η must lie on a geodesic $\mu_s : \mathbb{I} \to \mathfrak{U}_p(\mathbb{R})$. The proof relies on showing that $\mu_i(t+1) \notin \{\mu_i(t), \mu_j(t)\}$ when $w_{ii}, w_{ij} \in (0,1)$. The first term

$$w_{ii}(t) \left(\ell_p(\eta, \mu_i(t))^p - \ell_p(\eta, \mu_j(t))^p\right)$$

is strictly negative at $\eta = \mu_i(t)$ and strictly increasing as η moves from $\mu_i(t)$ to $\mu_j(t)$ and conversely $\ell_p(\eta, \mu_j(t))^p$ is strictly positive at $\eta = \mu_i(t)$ and strictly decreasing to zero as η moves from $\mu_i(t)$ to $\mu_j(t)$. Then for any $w_{ii} \in (0,1)$ and because ℓ_p is continuous it follows that there exists some μ_ε on μ_s with $\varepsilon > 0$ such that

$$w_{ii}(t) \left(\ell_p(\eta, \mu_i(t))^p - \ell_p(\eta, \mu_j(t))^p\right) \quad < \quad 0$$
$$\left|w_{ii}(t) \left(\ell_p(\eta, \mu_i(t))^p - \ell_p(\eta, \mu_j(t))^p\right)\right| \quad < \quad \ell_p(\eta, \mu_j(t))^p$$

on $\eta \in \mu_s$, $s \in [0, \varepsilon]$. Consequently, $\mu_i(t+1)$ is strictly decreasing on $\eta \in \mu_s$, $s \in [0, \varepsilon]$. Hence for any $w_{i1} \in (0,1)$ the point $\mu_i(t)$ cannot be a minimum. The same argument applies to $\mu_j(t)$. $\square$

The following is a simple consequence of the preceding result.

Corollary 1. *Consider the convex hull* $\mathrm{co}(\{\mu_i(0)\})$ *of all initial agent states in* $(\mathfrak{U}_p(\mathbb{R}), \ell_p)$*. If each agent applies (1) it follows that* $\mathrm{co}(\{\mu_i(t)\}) \subseteq \mathrm{co}(\{\mu_i(0)\})$ *for all* t*.*

The next result concerns an importance special case of the main result.

Lemma 4. *Suppose* $\mathcal{G}(\mathcal{V}, \mathcal{E})$ *is time-invariant and connected. Suppose the state of each agent is* $\mu_i(t) \in \mathfrak{U}_p(\mathbb{R})$ *and that each agent applies (1). Then there exists* $\mu^* \in \mathfrak{U}_p(\mathbb{R})$ *such that for any* $i \in \mathcal{V}$ *it holds that*

$$\lim_{t \to \infty} \ell_p(\mu_i(t), \mu^*) = 0.$$

Proof. It almost goes without saying that $\ell_2(\mu_i(t), \mu_j(t))^p = 0$, $\forall i, j \in \mathcal{V}$ with a constant $\mu_i(t)$ in $\mathfrak{U}_p(\mathbb{R})$ is an equilibrium state of (1). Consider a Lyapunov-like function $\nu(\mu) : \mathfrak{U}_p^n(\mathbb{R}) \to \mathbb{R}$ given by

$$\nu(\mu) = \sup_{\eta, \chi \in \{\mu_i(t)\}_{i \in \mathcal{V}}} \ell_p(\eta, \chi)^p \tag{3}$$

and note that $\nu(\mu) \geq 0$ with $\nu(\mu) = 0$ if and only if $\mu_i = \mu_j$ for all $i, j \in \mathcal{V}$ [4]. By Corollary 1 it follows that $\nu(\mu)$ is non-increasing along trajectories of (1). It suffices to show $\nu(\mu(t + n - 1)) < \nu(\mu(t))$ for each t. Firstly, pick a $t_0 \geq 0$ and note $\mathrm{co}(\{\mu_i(t_0)\}) \subseteq \mathrm{co}(\{\mu_i(0)\})$ and $f_{\ell_p, d}(\mathrm{co}(\{\mu_i(t_0)\})) \subseteq f_{\ell_p, d}(\mathrm{co}(\{\mu_i(0)\}))$ from Corollary 1 and where $f_{\ell_p, d}$ is an isometry given by (2). Without loss of generality, via Lemma 2, suppose that $f_{\ell_p, d}(\mathrm{co}(\{\mu_i(t_0)\}))$ is a l-sided polygon in $\mathbb{R}^2$ with $2 \leq l \leq |\mathcal{V}|$ on the collection of vertices $\{\mathbf{x}_j(t_0)\}$, $j \in \{1, \dots, l\}$ with

---

[3] This is because Fréchet averages in Euclidean space (on a set of input points) are associative, and can be found iteratively by computing the average initially for a pair of points (in a larger set of inputs), and then computing the average between this result and the next point (in the input set), and so on (adjusting the weights defining the average each time), see [47]. Owing to Lemma 2, this associativity property still holds here.

[4] We abuse notation here slightly. We use the shorthand μ as the argument in $\nu(\mu)$ to represent the collection of all agent measures $\{\mu_i(t)\}_{i \in \mathcal{V}}$ with $|\mathcal{V}| = n$.

$\mathbf{x}_j(t_0) \in \mathbb{R}^2$. If we chose a t_0 such that $l = 1$ then we would be done. Define the following set-valued function

$$h_j(t) = \left\{ i \in \mathcal{V} : f_{\ell_p,d}(\mu_i(t)) = \mathbf{x}_j(t_0) \right\}, \quad \forall j \in \{1, \dots, l\} \tag{4}$$

for each time $t \geq t_0$. It is immediate from Lemma 3 that $h_j(t+1) \subseteq h_j(t)$ for all $j \in \{1, \dots, l\}$; i.e. more generally, no agent state $f_{\ell_p,d}(\mu_i(t))$ which is not on the boundary of the l-sided polygon at time t can ever reach this same boundary at $t+1$ as a consequence of Lemma 3. Note that $|h_j(t_0)| \leq n - 1$ for all $j \in \{1, \dots, l\}$ with $l \geq 2$ at t_0. Recall the neighbour set at agent i is given by $\mathcal{N}_i(t)$. Because the network is connected, for each $k \in h_j(t_0)$ the neighbour set obeys $\mathcal{N}_k(t_0) \neq \emptyset$ for each $j \in \{1, \dots, l\}$. Then by Lemma 3 it follows that $h_j(t_0 + 1) \subset h_j(t_0)$ since at least one $k \in h_j(t_0)$ must be connected to an agent outside $h_j(t_0)$ and this agent's state must change $\mu_k(t_0) \neq \mu_k(t_0 + 1)$ as a consequence of Lemma 3 such that $f_{\ell_p,d}(\mu_k(t_0 + 1)) \neq \mathbf{x}_j$. At the next time $t_0 + 1$ it holds again that for each $k \in h_j(t_0 + 1)$ (assuming $h_j(t_0 + 1) \neq \emptyset$) the neighbour set obeys $\mathcal{N}_k(t_0 + 1) \neq \emptyset$ for each $j \in \{1, \dots, l\}$. Then by application of Lemma 3 it follows again that $h_j(t_0 + 2) \subset h_j(t_0 + 1) \subset h_j(t_0)$. Thus, $h_j(t+1) \subset h_j(t)$ is a strictly decreasing set-valued function unless $h_j(t) = \emptyset$. By at most time $t_0 + n - 1$ it follows that $h_j(t_0 + n - 1) = \emptyset$ and the argument can reset by redefining t_0. It follows that $f_{\ell_p,d}(\mathrm{co}(\{\mu_i(t_0 + n - 1)\})) \subset f_{\ell_p,d}(\mathrm{co}(\{\mu_i(t_0)\}))$ for all $t_0 \geq 0$. Following the proof of Lemma 2 we know $\mathrm{co}(\{\mu_i(t_0 + n - 1)\}) \subset \mathrm{co}(\{\mu_i(t_0)\})$ and thus because we chose t_0 arbitrarily $\nu(\mu(t + n - 1)) < \nu(\mu(t))$ for each $t \in \mathbb{N}$ unless $\mu_i(t + n - 1) = \mu_i(t)$, $\forall i$, as desired. The existence of a strictly decreasing Lyapunov function completes the proof. $\qquad \square$

The preceding lemma specialises the Theorem to the case where the network topology is connected and time-invariant (but otherwise arbitrary). This lemma is of interest on its own in many applications in which the topology is static. Proof of this lemma, given Lemmas 1-3, follows roughly the analysis of [5] on nonlinear consensus in the usual Euclidean metric space.

We are now ready to prove Theorem 1.

Proof. (**of Theorem 1**) The proof here relies on extending the previous lemma to the case where $\mathcal{G}(t)(\mathcal{V}, \mathcal{E}(t))$ is time-varying and for all $t_0 \in \mathbb{N}$ the graph union $\mathfrak{G}(t_0, \infty)$ is connected. Recall the same Lyapunov function (3) as used in the proof of Lemma 4 (we assume familiarity with the proof of Lemma 4 going forward).

We note that it suffices to show that there is a countably infinite number of finite time intervals $t \in [t_0^q, \widehat{t_0^q}]$, $q \in \mathbb{N}$ such that $\nu(\mu(t_0^q + \widehat{t_0^q})) < \nu(\mu(t_0^q))$.

Pick $t_0^q \geq 0$, $q \in \mathbb{N}$ so $f_{\ell_p,d}(\mathrm{co}(\{\mu_i(t_0^q)\}))$ is a l-sided polygon in $\mathbb{R}^2$ with $2 \leq l \leq |\mathcal{V}|$ on the collection of vertices $\{\mathbf{x}_j(t_0^q)\}$, $j \in \{1, \dots, l\}$ with $\mathbf{x}_j(t_0^q) \in \mathbb{R}^2$. Recall (4). Then define a sequence of times $\{t_{s(j)}^q\}$, $s(j) \in \mathbb{N}$ each greater than t_0^q for each $j \in \{1, \dots, l\}$ with $l \geq 2$. The connectivity condition implies the existence of such a sequence for each j with the property that, if $h_j(t_{s(j)}^q) \neq \emptyset$, there exists a $k \in h_j(t_{s(j)}^q)$ that is connected to an agent outside $h_j(t_{s(j)}^q)$. Then, this agent's state must change $\mu_k(t_{s(j)}^q) \neq \mu_k(t_{s(j)}^q + 1)$ as a consequence of Lemma 3 and $f_{\ell_p,d}(\mu_k(t_{s(j)}^q + 1)) \neq \mathbf{x}_j(t_0^q)$. Then $h_j(t_{s(j)}^q + 1) \subset h_j(t_{s(j)}^q)$ for all $j \in \{1, \dots, l\}$ unless obviously $h_j(t_{s(j)}^q) = \emptyset$. As in the proof of Lemma 4 it holds that $s(j) \geq n - 1$ implies $h_j(t_{s(j)+1}^q) = \emptyset$ for all j. Let $\widehat{t_0^q} = \min\{t \in \mathbb{N} : t > t_0^q, \ s(j) \geq n - 1, \forall j\}$ and note then that the interval $t \in [t_0^q, \widehat{t_0^q}]$ is finite owing to the connectivity condition. Moreover, as in the proof of Lemma 4 one can then show that $\nu(\mu(\widehat{t_0^q})) < \nu(\mu(t_0^q))$. Restart the argument by picking t_0^{q+1} to be equal or sufficiently close to $\widehat{t_0^q}$ and note that the connectivity condition then implies the number of such (finite) intervals $t \in [t_0^q, \widehat{t_0^q}]$ is countably infinite on $q \in \mathbb{N}$.

We thus have a strictly decreasing Lyapunov function $\nu(\mu(\widehat{t_0^q})) < \nu(\mu(t_0^q))$ on the sequence of finite intervals $t \in [t_0^q, \widehat{t_0^q}]$, $q \in \mathbb{N}$ and this completes the proof. $\qquad \square$

3 Special Cases and Convergence Details

Firstly, given Theorem 1, it is worth noting the following result.

Proposition 1. *Consider a group of agents $\mathcal{V}$ and network $\mathcal{G}(t)(\mathcal{V}, \mathcal{E}(t))$ where each agent i has initial state $\mu_i(0) \in \mathfrak{U}_p(\mathbb{R})$ and updates its state $\mu_i(t)$ according to (1). Suppose for all $t_0 \in \mathbb{N}$ the graph union $\mathfrak{G}(t_0, \infty)$ is connected so that Theorem 1 applies and there exists $\mu^* \in \mathfrak{U}_p(\mathbb{R})$ such that $\lim_{t\to\infty} \ell_p(\mu_i(t), \mu^*) = 0$ for any $i \in \mathcal{V}$. Then there exists some symmetric weight matrix $\overline{\mathbf{W}} = [\overline{w}_{ij}] \in \mathbb{R}^{n \times n}$ with $\overline{w}_{ij} \in (0,1)$ and $\sum_{j \in \mathcal{V}} \overline{w}_{ij} = 1$ for all i such that for any $i \in \mathcal{V}$*

$$\mu^* = \operatorname*{argmin}_{\eta \in \mathfrak{U}_p(\mathbb{R})} \sum_{j \in \mathcal{V}} \overline{w}_{ij} \ell_p(\eta, \mu_j(0))^p$$

where we emphasize that $\overline{\mathbf{W}}$ is not (generally) the same as $\mathbf{W}(t)$ but it is solely dependent on the sequence $\mathbf{W}(t)$, $t \in \mathbb{N}$ and (possibly) the initial measures $\{\mu_i(0)\}_{i \in \mathcal{V}}$.

Proof of this proposition is straightforward given the actual convergence result stated in Theorem 1. This result states that the common measure which all agent states converge to is within the convex hull of all initial agent measures in $\mathfrak{U}_p$.

An interesting open problem is how one can design the evolution of $\mathbf{W}(t)$, $t \in \mathbb{N}$ such that for a set of measures $\{\mu_i(0)\}_{i \in \mathcal{V}}$ the final weighting matrix $\overline{\mathbf{W}}$ specifies a limit μ^*, i.e. $\lim_{t\to\infty} \ell_p(\mu_i(t), \mu^*) = 0$ for any $i \in \mathcal{V}$, that is optimal, or desired, in some sense (e.g. minimum variance over all possible $\overline{\mathbf{W}}$ given $\mu_i(0) \in \mathfrak{U}_p(\mathbb{R})$, $i \in \mathcal{V}$).

In the remainder of this section we consider convergence to particular limits of interest, e.g. to a limit equally close to all agent's initial measures. We also consider convergence speeds and we consider computational aspects of the update protocol for given classes of input measures.

3.1 Convergence with Gaussian Measures

In this subsection we consider the evolution of the operation (1) at each agent $i \in \mathcal{V}$ when $\mu_i(0)$ is a Gaussian measure. We consider the case in which $p = 2$ and firstly consider measures on $\mathfrak{U}_2(\mathbb{R}^m)$ for which the Wasserstein metric definition can be straightforwardly extended.

Lemma 5. *[41, 42] Suppose that $\mu_i(t) \in \mathfrak{U}_2(\mathbb{R}^m)$ for all $i \in \mathcal{V}$ admits a Gaussian probability density of the form $\mathcal{N}(\mathbf{p}_i, \mathbf{P}_i)$. Then the solution to*

$$\mu_i(t+1) = \operatorname*{argmin}_{\eta \in \mathfrak{U}_2(\mathbb{R}^m)} \sum_{j \in \mathcal{N}_i(t)} w_{ij}(t) \ell_2(\eta, \mu_j(t))^2$$

is the Gaussian measure $\mu_i(t+1) \in \mathfrak{U}_2(\mathbb{R}^m)$ of density $\mathcal{N}(\mathbf{q}, \mathbf{Q})$ where $\mathbf{q} = \sum_{j \in \mathcal{N}_i(t)} w_{ij}(t) \mathbf{p}_j$ and $\mathbf{Q}$ is the unique positive-definite solution to

$$\mathbf{Q} = \sum_{j \in \mathcal{N}_i(t)} w_{ij}(t) \left(\mathbf{Q}^{1/2} \mathbf{P}_j \mathbf{Q}^{1/2} \right)^{1/2}$$

which is guaranteed to exist.

In the scalar case it follows that $\mu_i(t+1) \in \mathfrak{U}_2(\mathbb{R})$ is a Gaussian measure of density $\mathcal{N}(q, Q)$ where the variance is given by

$$Q = \left(\sum_{j \in \mathcal{N}_i(t)} w_{ij}(t) P_j^{1/2} \right)^2$$

and the updated mean q is given as in Lemma 5, i.e. $q = \sum_{j \in \mathcal{N}_i(t)} w_{ij}(t) p_j$.

Proposition 2. *Consider a group of agents $\mathcal{V}$ and a connected time-invariant network $\mathcal{G}(\mathcal{V}, \mathcal{E})$. Assume $\mathbf{W}$ is doubly stochastic and that $\mu_i(0) \in \mathfrak{U}_2(\mathbb{R})$ admits a Gaussian density $\mathcal{N}(p_i(0), P_i(0))$. Then $\mu_i(t+1) \in \mathfrak{U}_2(\mathbb{R})$ in (1) admits a Gaussian density $\mathcal{N}(p_i(t+1), P_i(t+1))$ where*

$$p_i(t+1) = \sum_{j \in \mathcal{N}_i} w_{ij} p_j(t),$$
$$P_i^{1/2}(t+1) = \sum_{j \in \mathcal{N}_i} w_{ij} P_j^{1/2}(t).$$

Moreover we have for any $i \in \mathcal{V}$ $\lim_{t \to \infty} \ell_2(\mu_i(t), \mu^*) = 0$ exponentially fast where μ^* satisfies

$$\mu^* = \frac{1}{n} \underset{\eta \in \mathfrak{U}_2(\mathbb{R})}{\operatorname{argmin}} \sum_{j \in \mathcal{V}} \ell_2(\eta, \mu_j(0))^2.$$

Proof. With Gaussian initial measures the operation (1) collapses to the standard linear consensus update (on the mean and variance) as shown in Lemma 5. Moreover, linear consensus in $\mathbb{R}$ over a time-invariant network with a doubly stochastic weighting matrix leads asymptotically to 'average' consensus [9]. Thus, in this case, the mean and standard deviation converge to the averages of all initial agent values and the result follows. $\qquad\square$

We have introduced a special scenario in the above proposition involving time-invariant networks and then provided a particular (though popular) scalar average consensus algorithm [7] to update the mean and variance (standard deviation) at each iteration. More general results on average consensus that allow for time-varying networks, finite-time convergence etc. [4, 6, 7, 19] may be substituted (but we do not explore this topic further here).

Although the update (1) is linear (in mean/variance) and closed in the event of Gaussian input measures, this is not generally the case, and the consensus problem (1) is, in general, inherently nonlinear. Nevertheless, we subsequently show that certain convergence properties of (1) at each agent follow by analysing a linear consensus protocol.

3.2 General Convergence Speeds and Average Wasserstein Consensus

In this subsection we consider only undirected, connected, and time-invariant network graphs $\mathcal{G}(\mathcal{V}, \mathcal{E})$. We consider only measures with finite second moment and we work solely in the Wasserstein metric space denoted by $(\mathfrak{U}_2(\mathbb{R}), \ell_2)$.

The first result considers the convergence speed of the entire group of agents under the protocol (1).

Proposition 3. *Consider a group of agents $\mathcal{V}$ and a connected time-invariant network $\mathcal{G}(\mathcal{V}, \mathcal{E})$ where each agent i updates its state $\mu_i(t) \in \mathfrak{U}_2(\mathbb{R})$ according to (1). Then there exists $\mu^* \in \mathfrak{U}_2(\mathbb{R})$ such that for any $i \in \mathcal{V}$,*

$$\lim_{t \to \infty} \ell_2(\mu_i(t), \mu^*) = 0$$

at an exponential rate.

Proof. If $\mu_i(0) \in \mathfrak{U}_2(\mathbb{R})$ then the solution to (1) at any $i \in \mathcal{V}$ and any $t \in \mathbb{N}$ can be written in the form

$$\mu_i(t+1)(M) = \left(\sum_{j \in \mathcal{N}_i} w_{ij} T_j^i(t) \right) \# \mu_i(t)(M)$$

for all Borel sets M on $(\mathbb{R}, d)$; see [42]. Here $(T_j^i(t)) \# \mu_i(t)$ denotes the push forward of $\mu_i(t)$ to $\mu_j(t)$ through the non-decreasing measurable map $T_j^i(t) : \mathbb{R} \to \mathbb{R}$ such that $(T_j^i(t)) \# \mu_i(t) = \mu_j(t)$. Obviously, we have $(T_i^i(t)) \# \mu_i(t) = \mu_i(t)$. For any measure $\psi(t)$ dominated by the Lebesgue measure on $\mathbb{R}$ it follows that

$$\left(\sum_{j \in \mathcal{N}_i} w_{ij} T_j^i(t) \right) \# \mu_i(t)(M) = \left(\sum_{j \in \mathcal{N}_i} w_{ij} T_j(t) \right) \# \psi(t)(M)$$

where $(T_j(t))\#\psi(t)$ denotes the push forward of $\psi(t)$ to $\mu_j(t)$; see [48]. Write the cumulative distribution function for each $\mu_i(t) \in \mathfrak{U}_2(\mathbb{R})$ by $F_i(x) : \mathbb{R} \to \mathbb{R}$ and $F_i(x) = \mu_i((-\infty, x])$. Define its inverse $F_i^+(x) : \mathbb{R} \to \mathbb{R}$ by

$$F_i^+(x) = \inf_y \{y \in \mathbb{R} : F_i(x) \geq y\}$$

for all $x \in \mathbb{R}$. One can show [42] that $T_j^i(t) = F_j^+ \circ F_i$ or if $\psi(t)$ is uniform on $[0, 1]$ then $T_j(t) = F_j^+$ and

$$\mu_i(t+1)(M) = \left(\sum_{j \in \mathcal{N}_i} w_{ij} F_j^+\right) \# \psi(t)(M)$$

with $\psi(t)$ uniform on $[0, 1]$. It follows directly that the solution to (1) at any $i \in \mathcal{V}$ and any $t \in \mathbb{N}$ has an inverse cumulative distribution function given by

$$F_i^+(t+1)(x) = \sum_{j \in \mathcal{N}_i} w_{ij} F_j^+(t)(x) \tag{5}$$

for all $x \in \mathbb{R}$.

Now one can stack these functions so $F^+(t+1)(\mathbf{x}) = \mathbf{W} F^+(t)(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$. From the assumed network connectivity condition and the weighting assumptions we conclude that $\mathbf{W}$ is row-stochastic and primitive with a distinct maximum eigenvalue of 1. The remaining $n-1$ eigenvalues have an absolute value strictly less than 1. The convergence rate of $F^+(t)(\mathbf{x})$ is determined by the convergence rate of $\mathbf{W}^t$ to the rank one matrix $\mathbf{1}\mathbf{u}^\top$ associated with the maximum eigenvalue. Writing

$$\mathbf{W}^t = \sum_{i=1}^n \lambda_i^t \mathbf{v}_i \mathbf{u}_i^\top = \mathbf{1}\mathbf{u}^\top + \sum_{i=2}^n \lambda_i^t \mathbf{v}_i \mathbf{u}_i^\top$$

where λ_i is the i'th eigenvalue of $\mathbf{W}$, it then follows that $\|\mathbf{W}^t - \mathbf{1}\mathbf{u}^\top\| = \|\sum_{i=2}^n \lambda_i^t \mathbf{v}_i \mathbf{u}_i^\top\|$ vanishes exponentially at a rate dominated by the absolute value of the second largest eigenvalue (which is strictly less than 1) and the proof is complete. $\qquad\square$

Note that a time-invariant network model is certainly not necessary for exponential convergence but we do not consider further generalisation in this work. It is important to note that the time-varying network connectivity condition allowed in Theorem 1 is also certainly too weak to ensure exponential convergence in general. Indeed, Theorem 1 does not even require the network to be jointly connected until some arbitrary finite future time.

Corollary 2. *Consider a group of agents $\mathcal{V}$ and a connected time-invariant network $\mathcal{G}(\mathcal{V}, \mathcal{E})$. Suppose that $\mathbf{W}$ is doubly stochastic and that each agent i updates its state $\mu_i(t) \in \mathfrak{U}_2(\mathbb{R})$ according to (1). For any $i \in \mathcal{V}$, we have $\lim_{t\to\infty} \ell_2(\mu_i(t), \mu^*) = 0$ at an exponential rate where*

$$\mu^* = \frac{1}{n} \inf_{\eta \in \mathfrak{U}_2(\mathbb{R})} \sum_{j \in \mathcal{V}} \ell_2(\eta, \mu_j(0))^2.$$

Proof. This result follows again because linear consensus in $\mathbb{R}$ over a time-invariant network with a doubly stochastic weighting matrix leads asymptotically to 'average' consensus [9]. Looking at (5) we see that (nonlinear) consensus via (1) is related to (linear) consensus in the space of inverse cumulative distribution functions. Moving from the limiting inverse cumulative distribution function to a probability measure does not change the limiting $1/n$ averaging coefficient. $\qquad\square$

This corollary provides sufficient conditions[5] under which exponential convergence to a measure is achieved and where the consensus measure achieved asymptotically at each agent is an average distance to all initial measures. We note that other consensus measures may be more desirable, e.g. one may want to reach an agreement on that measure with the smallest variance within the convex hull of all initial measures.

---

[5] A time-invariant network topology and a doubly stochastic weighting matrix. The time-invariance constraint can be relaxed (it is just sufficient) but we do not consider generalisation here.

3.3 Computational Aspects of the Update Protocol

In the case of Gaussian input measures, we have shown that the updating step of our consensus algorithm can be performed in closed form and that the resulting algorithm resembles a particular case of standard linear consensus in $\mathbb{R}$, e.g. see [9]. In the general case, the optimization problem (1) typically does not admit such a closed form solution. However, it is convex [42, 44] and thus numerical methods are feasible and already exist in a number of cases; see [43, 44, 49].

Moreover, consider the important scenario where all the input initial measures are (weighted) empirical measures

$$\mu_i^N(0)\,(dx) = \sum_{j=1}^{N} \alpha_j^i \delta_{x_j^i(0)}\,(dx)\,,$$

where $\delta_y\,(dx)$ denotes the delta-Dirac measure located at y, $\alpha_j^i \geq 0$ and $\sum_{j=1}^{N} \alpha_j^i = 1$. In this case, the minimization in (1) can be solved exactly via a finite-dimensional linear program [50, 51] and the resulting measure is also an empirical measure[6]. However, the computational requirements of this linear program may explode quickly with the number of input measures and the number of atoms of these measures; see [33, 34, 50, 51]. Nevertheless, numerous fast approximation methods have been derived [44, 47, 49, 52]. The details of these algorithms are beyond the scope of this work, but it follows that operation (1) is thus practical in the application important case involving empirical measures[7].

Consider, more generally, arbitrary input measures on $\mathbb{R}$. Convexity of the minimisation problem is advantageous in general, but in this case (i.e. with measures on the line) there are yet further virtues. Refer to (5) which is intimately related to (1). The update in (5) is typically computable in closed-form and thus we 'almost' have a general closed-form expression for (1) already. This relationship between the update (1) and the inverse cumulative distribution function has been explored in [42–44] with example computations and as a lead into more general computational results. We refer the interested reader to this literature. We also note in conclusion that the computational aspect of this update equation is an ongoing research topic.

4 Discussion and Applications

The output of operation (1) is known as the Wasserstein barycenter in the literature and the limit to which all agents convergence $\mu^* \in \mathfrak{U}_p(\mathbb{R})$ is similarly a Wasserstein barycentre (on all initial agent's measures). In other words, this work studies the convergence properties of a consensus algorithm concerned with distributed (iterative) computation of the Wasserstein barycentre over a (possibly) time-varying, arbitrary, network topology.

While this is the first such study in this direction, potential applications/uses for the Wasserstein barycentre (itself) have been considered previously in a number of fields [33, 34, 44, 47, 49, 51, 52, 52–54] and this list is by no means exhaustive.

Arguably the most popular domain in which the Wasserstein barycentre has found applications is in computer vision and image/video processing [44, 51, 52]. We do not consider specifics here but the interested reader may consult [44] where numerous examples and a detailed discussion is given. It is noted [44] that state-of-the-art advancements in a number of related problems have arisen via the use of Wasserstein barycenters. Importantly, both Gaussian and discrete measures find applicability through the Wasserstein barycentre in computer vision and image/video processing; again see [44].

---

[6]As discussed previously, for special, certainly non-generic, arrangements of discrete measures the minimisation in (1) may not have a unique solution (though a solution always exists) [44]. We ignore these degenerate special cases (leading to non-uniqueness) since they may only arise under constraints of little practical interest.

[7]Interestingly, if each input measure is defined by a single Dirac (in $(\mathfrak{U}_2(\mathbb{R}), \ell_2)$, i.e. with $p = 2$), then the classical (linear) consensus algorithm in $\mathbb{R}$ is recovered, e.g. as in [5, 9]. Of course, typically one is interested in more general empirical input measures.

Applications in machine learning and Bayesian statistics have also made use of the Wasserstein barycentre [43, 49, 53] and it is envisioned that this technology (and the related optimal transportation problem) will find wider adoption in this field. In this setting, distributed (or even parallel) computation of the Wasserstein barycentre is likely important; e.g. distributed Bayesian computation on large data sets is the subject of [53].

Information fusion and distributed estimation (e.g distributed particle filtering) have been studied in the context of Wasserstein barycenters [33, 34, 43]. This topic is one that may use directly the consensus algorithm proposed [33] (with empirical measures as inputs; e.g. from the output of Monte Carlo estimators like particle filters). Data fusion with correlated Gaussian measures has been shown to be consistent (i.e. not optimistic) with this algorithm when the correlation is ignored [34].

5   Concluding Remarks

Distributed consensus in the Wasserstein metric space of probability measures was introduced in this paper. It is shown that convergence of the individual agents' measures to a common measure value is guaranteed if a relatively weak network connectivity condition is satisfied. The measure that is achieved asymptotically at each agent is the measure that minimises a weighted sum of its Wasserstein distances to these initial measures and is known as the Wasserstein barycentre in the literature.

Finally, we note that following [5], it would be straightforward to consider an extension to the case in which the network topology is directed and one expects analogous results (concerning connectivity) to apply in the Wasserstein space considered herein. For brevity, and notational simplicity, we do not explore this scenario further. Moreover, one may seek analogous results in the Wasserstein metric space of measures defined on the Borel sets of $(\mathbb{R}^m, d)$ for some $m \geq 2$. We conjecture that similar results hold in this case. However while many of the lemmas used herein carry over immediately, this generalization is not straightforward. Indeed, the Wasserstein metric space in such cases is positively curved so it does not resemble Euclidean space at all and it is not even CAT(0).

References

[1] M.H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, March 1974.

[2] J.N. Tsitsiklis, D.P. Bertsekas, and M. Athans. Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Trans. on Automatic Control*, 31(9):803–812, September 1986.

[3] A. Jadbabaie, J. Lin, and A.S. Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions on Automatic Control*, 48(6):988–1001, June 2003.

[4] R. Olfati-Saber and R.M. Murray. Consensus problems in networks of agents with switching topology and time-delays. *IEEE Transactions on Automatic Control*, 49(9):1520–1533, September 2004.

[5] L. Moreau. Stability of multiagent systems with time-dependent communication links. *IEEE Transactions on Automatic Control*, 50(2):169–182, February 2005.

[6] W. Ren and R.W. Beard. Consensus seeking in multiagent systems under dynamically changing interaction topologies. *IEEE Transactions on Automatic Control*, 50(5):655–661, 2005.

[7] R. Olfati-Saber, J.A. Fax, and R.M. Murray. Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*, 95(1):215–223, January 2007.

[8] M. Cao, A.S. Morse, and B.D.O. Anderson. Reaching a consensus in a dynamically changing environment: A graphical approach. *SIAM Journal on Control and Optimization*, 47(2):575–600, 2008.

[9] L. Xiao, S. Boyd, and S. Lall. A scheme for robust distributed sensor fusion based on average consensus. In *Proc. of the 4th International Symposium Information Processing in Sensor Networks*, pages 63–70, Los Angeles, California, 2005.

[10] D.P. Spanos and R.M. Murray. Distributed sensor fusion using dynamic consensus. In *Proc. of the 16th IFAC World Congress*, Prague, Czech Republic, July 2005.

[11] R. Olfati-Saber. Distributed Kalman filtering for sensor networks. In *Proc. of the 46th IEEE Conference on Decision and Control*, pages 5492–5498, New Orleans, Louisiana, December 2007.

[12] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri. Distributed Kalman filtering based on consensus strategies. *IEEE Journal on Selected Areas in Communications*, 26(4):622–633, 2008.

[13] F.S. Cattivelli and A.H. Sayed. Diffusion strategies for distributed Kalman filtering and smoothing. *IEEE Transactions on Automatic Control*, 55(9):2069–2084, 2010.

[14] Q. Hui and W.M. Haddad. Distributed nonlinear control algorithms for network consensus. *Automatica*, 44(9):2375–2381, 2008.

[15] W. Yu, G. Chen, and M. Cao. Consensus in directed networks of agents with nonlinear dynamics. *IEEE Transactions on Automatic Control*, 56(6):1436–1441, 2011.

[16] A. Ajorlou, A. Momeni, and A.G. Aghdam. Sufficient conditions for the convergence of a class of nonlinear distributed consensus algorithms. *Automatica*, 47(3):625–629, 2011.

[17] J. Cortés. Distributed algorithms for reaching consensus on general functions. *Automatica*, 44(3):726–737, 2008.

[18] X. Wang and Y. Hong. Distributed finite-time χ-consensus algorithms for multi-agent systems with variable coupling topology. *Journal of Systems Science and Complexity*, 23(2):209–218, 2010.

[19] J. Cortés. Finite-time convergent gradient flows with applications to network consensus. *Automatica*, 42(11):1993–2000, 2006.

[20] M. Zhu and S. Martínez. Discrete-time dynamic average consensus. *Automatica*, 46(2):322–329, 2010.

[21] Y. Hong, J. Hu, and L. Gao. Tracking control for multi-agent consensus with an active leader and variable topology. *Automatica*, 42(7):1177–1182, 2006.

[22] S.H. Strogatz. From Kuramoto to Crawford: Exploring the onset of synchronization in populations of coupled oscillators. *Physica D: Nonlinear Phenomena*, 143(1):1–20, 2000.

[23] Z. Li, Z. Duan, G. Chen, and L. Huang. Consensus of multiagent systems and synchronization of complex networks: A unified viewpoint. *IEEE Trans. on Circuits and Systems I*, 57(1):213–224, 2010.

[24] S. Emre Tuna and R. Sepulchre. Consensus under general convexity. In *Proc. of the 46th IEEE Conference on Decision and Control*, pages 294–299, New Orleans, USA, 2007.

[25] A. Sarlette and R. Sepulchre. Consensus optimization on manifolds. *SIAM J. on Control and Optimization*, 48(1):56–76, 2009.

[26] I. Matei and J.S. Baras. The asymptotic consensus problem on convex metric spaces. In *Proc. of the 2nd IFAC Workshop on Distributed Estimation and Control in Networked Systems*, Annecy, France, Sept. 2010.

[27] R. Sepulchre. Consensus on nonlinear spaces. *Annual Reviews in Control*, 35(1):56–64, 2011.

[28] P. Grohs. Wolfowitz's theorem and convergence of consensus algorithms in hadamard spaces. Technical Report SAM Report 2012-27, ETH, Zurich, Switzerland, August 2012.

[29] S. Bonnabel. Stochastic gradient descent on Riemannian manifolds. *IEEE Transactions on Automatic Control*, 58(9):2217–2229, September 2013.

[30] C.R. Givens and R.M. Shortt. A class of Wasserstein metrics for probability distributions. *The Michigan Mathematical Journal*, 31(2):231–240, 1984.

[31] R. Olfati-Saber, E. Franco, E. Frazzoli, and J.S. Shamma. Belief consensus and distributed hypothesis testing in sensor networks. In *Proc. of the Workshop on Network Embedded Sensing and Control*, Notre Dame University, South Bend, Indiana, October 2005.

[32] G. Battistelli and L. Chisci. Kullback–Leibler average, consensus on probability densities, and distributed state estimation with guaranteed stability. *Automatica*, 50(3):707–718, 2014.

[33] A.N. Bishop. Gossip-based distributed data fusion of empirical probability measures. In *Proc. of the 2014 IEEE Workshop on Statistical Signal Processing*, pages 372–375, Gold Coast, Australia, 2014.

[34] A.N. Bishop. Information fusion via the Wasserstein barycenter in the space of probability measures: Direct fusion of empirical measures and Gaussian fusion with unknown correlation. In *Proc. of the 17th International Conference on Information Fusion*, Salamanca, Spain, July 2014, (submitted: February 18, 2014).

[35] I.L. Manuel and A.N. Bishop. Distributed Monte Carlo information fusion and distributed particle filtering. In *Proc. of the 19th IFAC World Congress*, pages 8681–8688, Cape Town, South Africa, August 2014.

[36] A.N. Bishop and A. Doucet. Distributed nonlinear consensus in the space of probability measures. In *Proc. of the 19th IFAC World Congress*, Cape Town, South Africa, August 2014, (submitted: October 27, 2013).

[37] C. Villani. *Topics in Optimal Transportation*. American Mathematical Society, 2003.

[38] L. Ambrosio, N. Gigli, and G. Savaraé. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Birkhauser Verlag, Basel, Switzerland, 2005.

[39] B. Kloeckner. A geometric study of Wasserstein spaces: Euclidean spaces. *Annali della Scuola Normale Superiore di Pisa*, IX(2):297–323, 2010.

[40] J. Bertrand and B. Kloeckner. A geometric study of Wasserstein spaces: Hadamard spaces. *Journal of Topology and Analysis*, 4(4):515–542, December 2012.

[41] R.J. McCann. A convexity principle for interacting gases. *Advances in Mathematics*, 128(1):153–179, 1997.

[42] M. Agueh and G. Carlier. Barycenters in the Wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, 2011.

[43] J. Bigot and T. Klein. Consistent estimation of a population barycenter in the Wasserstein space. *arXiv e-print arXiv:1212.2562*, 2012.

[44] N. Bonneel, J. Rabin, G. Peyré, and H. Pfister. Sliced and Radon Wasserstein barycenters of measures. *Journal of Mathematical Imaging and Vision*, 51(1):22–45, 2015.

[45] D. Burago, Y. Burago, and S. Ivanov. *A Course in Metric Geometry*. American Mathematical Society, Providence, R.I., 2001.

[46] M.R. Bridson and A. Haefliger. *Metric Spaces of Non-Positive Curvature*. Springer, Berlin, Germany, 1999.

[47] E. Boissard, T. Le Gouic, and J.-M. Loubes. Distribution's template estimate with Wasserstein metrics. *arXiv e-print arXiv:1111.5927*, 2011.

[48] N. Bonneel, J. Rabin, G. Peyré, and H. Pfister. Sliced and Radon Wasserstein barycenters of measures. *Journal of Mathematical Imaging and Vision*, pages 1–24, 2013.

[49] M. Cuturi and A. Doucet. Fast computation of Wasserstein barycenters. In *Proc. of the 31st International Conference on Machine Learning*, pages 685–693, Beijing, China, 2014.

[50] W. Gangbo and A. Swiech. Optimal maps for the multidimensional Monge-Kantorovich problem. *Communications on Pure and Applied Mathematics*, 51(1):23–45, January 1998.

[51] J. Rabin, G. Peyré, J. Delon, and M. Bernot. Wasserstein barycenter and its application to texture mixing. In *Scale Space and Variational Methods in Computer Vision*, pages 435–446. Springer, 2012.

[52] G. Carlier, A. Oberman, and E. Oudet. Numerical methods for matching for teams and Wasserstein barycenters. *arXiv e-print arXiv:1411.3602*, 2014.

[53] S. Srivastava, V. Cevher, Q. Tran-Dinh, and D.B. Dunson. WASP: Scalable Bayes via barycenters of subset posteriors. In *Proc. of the 18th International Conference on Artificial Intelligence and Statistics*, pages 912–920, San Diego, USA, 2015.

[54] J. Arroyo, G. González-Rivera, C. Maté, and A.M. San Roque. Smoothing methods for histogram-valued time series: An application to value-at-risk. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 4(2):216–228, April 2011.